

REM WORKING PAPER SERIES

Explainable models of credit losses

João A. Bastos, Sara M. Matos

REM Working Paper 0161-2021

February 2021

REM – Research in Economics and Mathematics

Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

Any opinions expressed are those of the authors and not those of REM. Short, up to two paragraphs can be cited provided that full credit is given to the authors.





REM – Research in Economics and Mathematics

Rua Miguel Lupi, 20
1249-078 LISBOA
Portugal

Telephone: +351 - 213 925 912

E-mail: rem@iseg.ulisboa.pt

<https://rem.rc.iseg.ulisboa.pt/>



<https://twitter.com/ResearchRem>

<https://www.linkedin.com/company/researchrem/>

<https://www.facebook.com/researchrem/>

Explainable models of credit losses

João A. Bastos*, Sara M. Matos†

Abstract

Credit risk management is an area where regulators expect banks to have transparent and auditable risk models, which would preclude the use of more accurate black-box models. Furthermore, the opaqueness of these models may hide unknown biases that may lead to unfair lending decisions. In this study, we show that banks do not have to sacrifice prediction accuracy at the cost of model transparency to be compliant with regulatory requirements. We illustrate this by showing that the predictions of credit losses given by a black-box model can be easily explained in terms of their inputs. Because black-box models are better at uncovering complex patterns in the data, banks should consider the determinants of credit losses suggested by these models in lending decisions and pricing of credit exposures.

JEL Classification: G21 · G33 · C14 · C52

Keywords: Credit risk · Loss given default · Recovery rates · Explainable machine learning · Forecasting

1 Introduction

Model interpretability may be defined as the degree to which a human can understand the cause of its outputs. Credit risk management is an area where regulators expect banks to have explainable and auditable risk models. The importance of transparent risk models emerged after the implementation of the Basel II agreement (BCBS, 2006). On the one hand, under the Pillar I of the Basel guidelines, banks were allowed to use their own estimates of credit risk factors, such as the probability of default and the expected loss given default, for the purpose of calculating regulatory capital. On the other hand, Pillar II gave supervisors greater authority to assess the consistency and soundness of the risk assessment methodologies developed internally

*ISEG, Lisbon School of Economics and Management, Universidade de Lisboa; REM/CEMAPRE. E-mail: jbastos@iseg.ulisboa.pt.

†Deloitte. E-mail: samatos@deloitte.pt.

This work was supported by the FCT – Fundação para a Ciência e a Tecnologia (grant number UIDB/05069/2020).

by the banks. This precluded the use of black-box models. Furthermore, the opaqueness of these models may hide unknown biases that may lead to unfair lending decisions, and banks are interested in understanding which factors drive a particular decision.

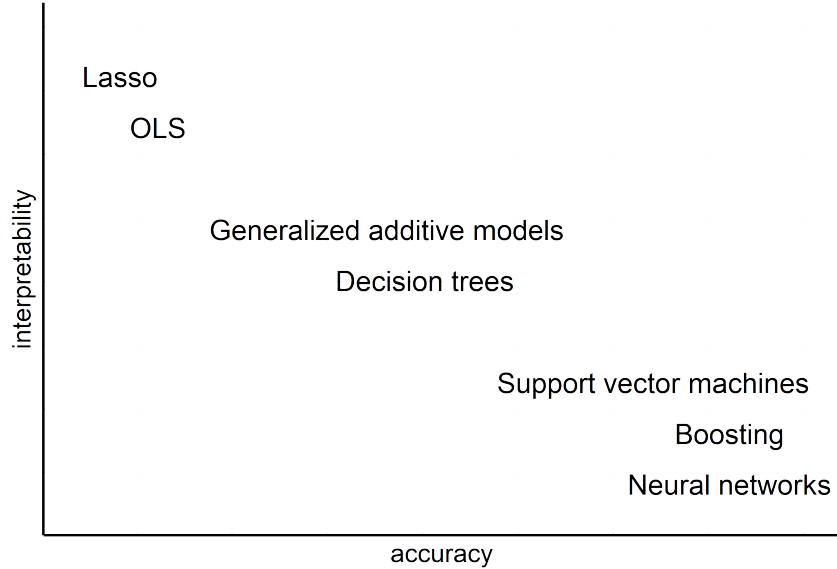


Figure 1: Trade-off between accuracy and interpretability. Easily explainable models are usually less accurate, whereas complex models are usually more accurate.

In statistical modeling, there is a trade-off between interpretability and prediction accuracy: models with outputs that are easily explainable in terms of their inputs are usually less accurate, whereas models that are complex and opaque have greater out-of-sample precision (James et al., 2014). Figure 1 provides an illustration of this trade-off. On the upper-left corner we have models such as Lasso (Tibshirani, 1996), and the workhorse of applied econometric analysis: the linear model estimated by least squares. These “glass-box” models have good inference properties and are easily explainable. However, due to their fixed functional form, they do not possess enough flexibility to fit complex data, and give poor accuracy on most tasks. On the center of the plot we have models such as decision trees (Breiman et al., 1983), and generalized additive models (Hastie and Tibshirani, 1990). These models are more flexible, and typically have better accuracy than linear models. With an additional effort we may understand which variables are driving the predictions. However, large decision trees are difficult to interpret, and additional techniques are required to understand which variables determine the model outcomes. On the lower-right corner, we have models that emerged from the machine learning literature, such as support vector machines (Vapnik, 1995), neural networks (Rumelhart et al., 1986), and boosting (Freund and Schapire, 1996). These models are highly flexible and have greater out-of-sample precision. Currently, deep neural networks are the state of the art in tasks such as image recognition, speech recognition, and natural language processing. Boosting methods were used in the majority of the winning solutions in data science competitions (Chen and Guestrin, 2016). However, these models are regarded as black-boxes. While this is not an issue when we

are only concerned with the performance of out-of-sample predictions, the opaqueness of these models is problematic when we must know which variables contributed to their outcomes.

The Basel agreements triggered a great interest in the modeling of loss given default (LGD) – the proportion of a debt exposure that the bank estimates to lose if the borrower is no longer able to comply with the contractual terms. The importance of LGD is not limited to the estimation of regulatory capital requirements, but is relevant to pricing credit instruments such as loans, bonds, and credit default swaps. Several studies on the determinants of LGD used linear models estimated by OLS (e.g., Acharya et al. (2007); Davidenko and Franks (2008)). Because credit losses are usually measured as a proportion of the exposure at default, and therefore limited to the interval $[0,1]$, this is not the most appropriate approach for modeling them. A better alternative is a fractional regression model estimated by quasi-maximum likelihood (Papke and Wooldridge, 1996). For instance, Dermine and Neto de Carvalho (2006) used this approach to model bank loan losses. Moody’s proprietary LossCalc v2 addresses the bounded nature of credit losses by transforming them via a beta distribution before applying a linear model (Gupton and Stein, 2005). Any of these models are glass-boxes that allow us to infer which variables are driving the predictions.

Bastos (2010) proposed using non-parametric decision trees to model bank loan losses. Since the estimated trees were small, credit losses could be easily explained as a set of if-then-else rules on the covariates. Furthermore, these models had better out-of-sample accuracy than parametric regressions. Bastos (2014) studied the performance of a “committee” of decision trees, in which each tree is estimated using bootstrap samples of the original data. This committee had better accuracy than a single decision tree or a parametric model. However, the interpretability vs. accuracy trade-off stepped in, and the simplicity of the if-then-else rules given by a single tree was lost. Other black-box models have also been considered to predict credit losses. For instance, Loterman et al. (2012) performed a large-scale benchmark study using 24 regression techniques that were evaluated on six real-life data sets obtained from major international banking institutions. They concluded that machine learning techniques, such as support vector machines and neural networks, have better performance in predicting credit losses. Again, these models are opaque and, by themselves, have limited value under the Pillar II of the Basel agreements.

In this study, we show that banks do not have to sacrifice predictive accuracy at the cost of model transparency to be compliant with the guidelines of the Basel agreements. In particular, we show that the predictions given by black-box models for credit losses given default can be interpreted in terms of their inputs. We can derive the relative importance of the regressors, and understand whether the relationships between credit losses and the covariates are positive or negative, linear or non-linear, convex or concave. Because black-box models perform substantially better than glass-box models at discovering complex structures in the data, banks should concentrate on the drivers of credit losses indicated by the former in their lending decisions and in the pricing of credit exposures. Therefore, banks can comply with the Basel guidelines while

pursuing the most effective allocation of capital requirements.

We will do so by comparing the drivers of credit losses given by three models:

1. Fractional regression model. This econometric model is adequate for modeling credit losses, since these are measured as a proportion of the outstanding debt at default. This parametric model is easily interpretable and sits on the upper left corner of Figure 1.
2. Decision trees. This is a non-parametric and mildly interpretable model that sits on the center of Figure 1. We can understand which variables are driving the predictions, since those are the ones participating in sequences of if-then-else conditions. The direction of the partial effects can be understood if the trees are small. We show that trees give better predictions than the fractional regression model.
3. Gradient boosting machine. This model consists of an *ensemble* of decision trees. While this is a very powerful technique for predicting credit losses, it consists of a black-box. This model sits on the lower right corner of Figure 1.

In the next section, we provide an overview of known risk drivers of credit losses using Moody’s Ultimate Recovery Database. In Section 3, we estimate the three aforementioned models. We show that the black-box model has much better out-of-sample and out-of-time accuracy than the others. In Section 4, we show how to interpret the black-box model outputs, and compare the resulting interpretations with those of the other models. Section 5 provides the conclusions.

2 Risk drivers of loss given default

Our data set is Moody’s Ultimate Recovery Database (URD), which describes US non-financial corporations holding over \$50 million in debt at the time of default. It contains 4630 defaulted instruments (bonds and loans) from 957 different obligors, covering default events between 1987 and 2010. In this database, credit losses are quantified by the “recovery rate”, which is simply $1 - \text{LGD}$. Moody’s URD includes three different valuation methods of the amount received by creditors at the resolution of default: settlement method, trading price method, and liquidity event method. It also indicates which of the methods Moody’s considers as the most representative of the actual recovery. We use the discounted recovery rate recommended by Moody’s. Figure 2 shows the distribution of discounted recovery rates. The distribution is bimodal with full recovery as the most common outcome. Approximately 20% of the debts recovered less than 10%, while 40% recovered more than 90%. The average discounted recovery rate is 59%.

2.1 Debt instrument type

Discounted recoveries depend strongly on the type of debt and its position in terms of claims priority. On a sample of over 700 defaulted bonds, Altman and Kishore (1996) find that the

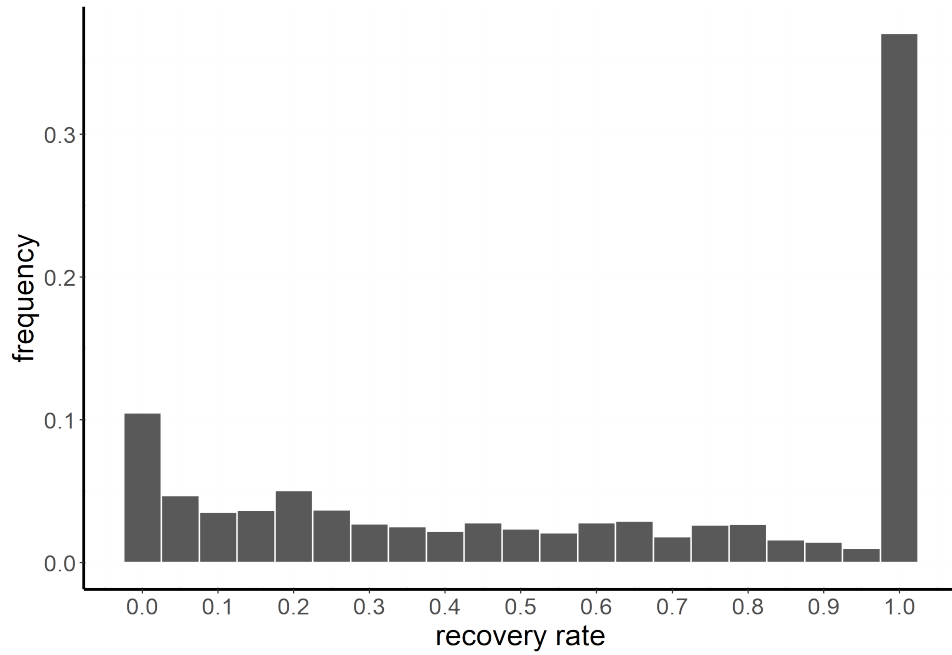


Figure 2: Distribution of discounted recovery rates from Moody's ultimate recovery database.

seniority of the bond does play an important role on the recovery. Figure 3 shows the discounted average recovery rate by type of exposure. Defaulted instruments can be separated into two groups: bonds (about 60% of the dataset) and loans. Bonds have an average recovery rate of 45%, while loans recover on average 80%, reflecting the typically higher credit position of loans in terms of claims priority. The average recovery rates by type of instrument vary between 19% (Junior subordinated bonds) and 85% (Revolver loans).

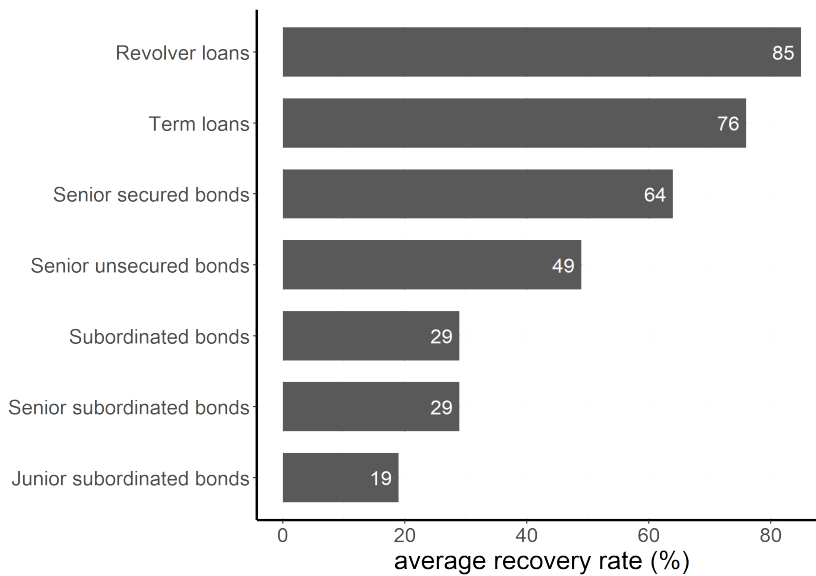


Figure 3: Average discounted recovery rate by debt instrument type.

2.2 Seniority within the liabilities of the firm

In our data set, firms had on average five different type of debts in their liability structure, and there is strong empirical evidence that the seniority of the debt within this structure has a substantial impact on recovery rates. For North American corporate debt issuers, Varma and Cantor (2005) conclude that seniority is one of the most important variables for determining the recovery rate. They show that the higher the value of the *debt cushion* (the value of debt junior to one’s own claim) the greater the amount that can be expected to recover. This result has economic plausibility since the larger the debt cushion, the greater the amount that is likely to be available for distribution to more senior applicants.

Average recovery as a function of			
	debt above	debt cushion	instrument amount
0.00 – 0.25	22%	93%	62%
0.25 – 0.50	36%	88%	54%
0.50 – 0.75	47%	70%	52%
0.75 – 1.00	71%	43%	47%

Table 1: Average discounted recovery rate by seniority of the debt within the liabilities of the firm. Debt above is the percentage of total liabilities senior to the debt. Debt cushion is the percentage of total liabilities junior to the debt. The last column shows the average recovery rate as a function of the instrument amount at default as a proportion of the firm’s total liabilities.

Table 1 shows that the amount received by creditors declines as the proportion of total liabilities senior to a given debt increases. Recovery rates average only 22% for those defaulted instruments with senior debt greater than 75% of total liabilities, compared to 71% for defaulted instruments with senior debt less than 25% of total liabilities. On the other hand, the average recovery rate increases with the percentage of total debt junior to an exposure. Recovery rates average 93% for defaulted instruments with junior debt greater than 75% of total liabilities, and 43% for those with junior debt less than 25% of total liabilities. The last column in Table 1 shows that the amount of an exposure at default relative to the firm’s total liabilities is associated to different recovery rates. Low relative amounts of an exposure at default have higher recoveries. On the other hand, recoveries decrease moderately as the relative instrument amount increases beyond 0.25.

Figure 4 shows the average recovery rate as a function of the rank of an exposure in terms of priority of claim in the liabilities of a firm. A ranking of 1 corresponds to the most senior debt, and 7 to the most junior. Naturally, recovery rates are substantially higher for those debts with the highest priority of claim. The relatively higher recovery for debts with rank 7 is a statistical artifact due to the small number of debts having this rank.

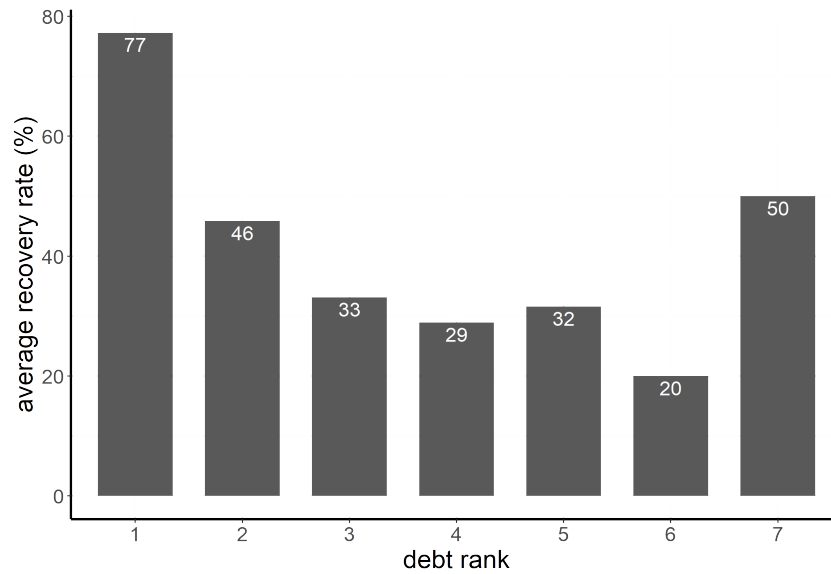


Figure 4: Average discounted recovery rate by debt rank.

2.3 Collateral

Credit losses also vary according to the existence and quality of collateral associated with the defaulted instruments. The average recovery rates by collateral type are shown in Figure 5. As expected, unsecured exposures have the lowest mean recovery rate. Debts secured by inventory accounts receivable and cash result in higher recoveries, which is also anticipated since these assets are easier to liquidate.

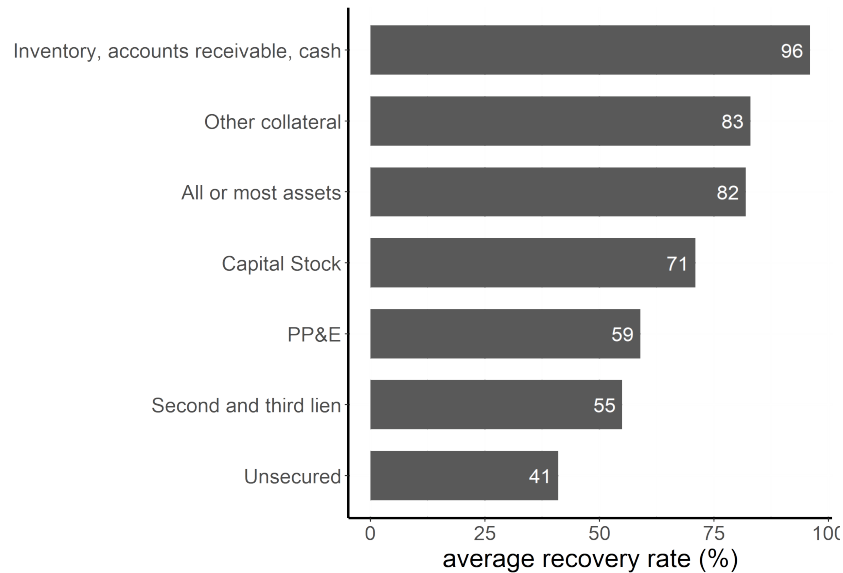


Figure 5: Average discounted recovery rate by collateral.

2.4 Industry sector

The counterparty industry sectors are one of the most frequently used explanatory variables in the estimation of credit losses. For corporate bonds, Altman and Kishore (1996) find evidence of similar recovery rates for a large number of industries, although great differences occur in a few sectors. We observe a similar pattern in Moody's URD. Figure 6 shows the sample mean recovery rate by Moody's industry classification. A large number of industries had recoveries around 60%. The Environment sector had a remarkably low average recovery rate of 29%, followed by the Telecommunication (42%) and Construction (48%) industries. On the opposite extreme, the Natural products and Energy industries had average recovery rates of 82% and 74%, respectively.

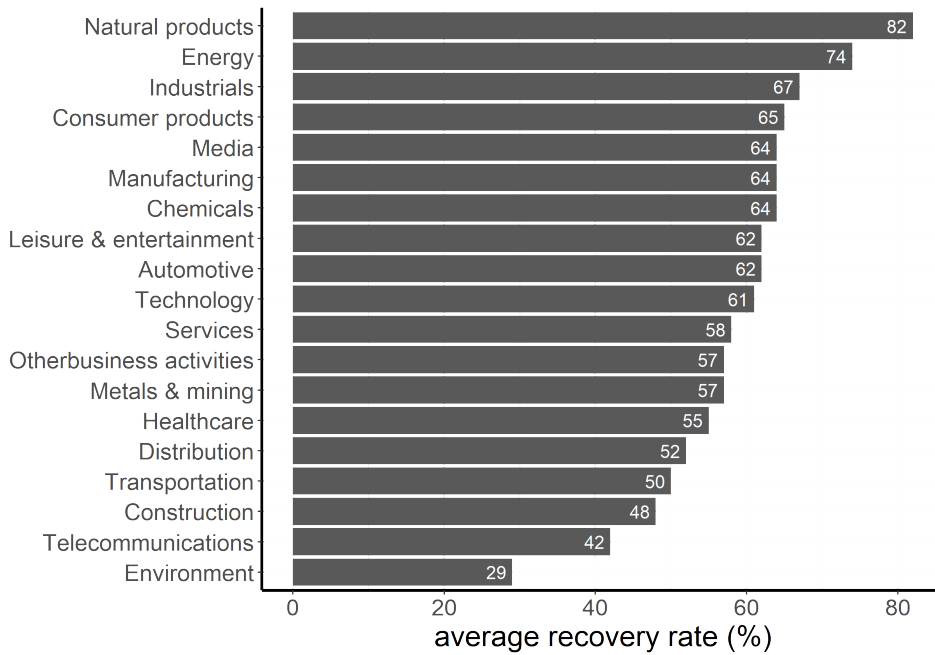


Figure 6: Average discounted recovery rate by industry sector.

3 Models

3.1 Fractional regression model

Our first model is a glass-box parametric model. Because recovery rates are bounded to $[0, 1]$ we want to estimate a parametric model suited for modeling fractional response variables. The model is

$$E(Y|\mathbf{X}) = G(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) = G(\mathbf{X}^T \boldsymbol{\beta}), \quad (1)$$

where $G(\cdot)$ satisfies $0 < G(z) < 1$ for all $z \in \mathbb{R}$. This condition ensures that the predicted values fall within the unit interval. An usual functional form for $G(\cdot)$ is the logistic function,

$$G(\mathbf{X}^T \boldsymbol{\beta}) = \frac{1}{1 + \exp(-\mathbf{X}^T \boldsymbol{\beta})}. \quad (2)$$

The non-linear estimation procedure consists of the maximization of the Bernoulli log-likelihood function (Papke and Wooldridge, 1996):

$$L(\boldsymbol{\beta}) \equiv \sum_{i=1}^n Y_i \log [G(\mathbf{X}_i^T \boldsymbol{\beta})] + (1 - Y_i) \log [1 - G(\mathbf{X}_i^T \boldsymbol{\beta})]. \quad (3)$$

The quasi-maximum likelihood estimator is consistent and asymptotically normal, regardless of the distribution of Y conditional on \mathbf{X} , and therefore the Z -statistics indicate which regressors are statistically significant in predicting recovery rates.

However, since the function $G(\cdot)$ is non-linear, the partial effects of the explanatory variables on the recovery rates are not constant. The partial effect of variable x_j on Y is

$$\frac{\partial E(Y|\mathbf{X})}{\partial x_j} = \frac{dG(\mathbf{X}^T \boldsymbol{\beta})}{d(\mathbf{X}^T \boldsymbol{\beta})} \beta_j. \quad (4)$$

Because $G(\mathbf{X}^T \boldsymbol{\beta})$ is strictly monotonic, the sign of the coefficient provides the direction of the partial effect. The contribution to the recovery rate of each variable can be obtained by estimating the sample average of the partial effects given by Equation (4).

We estimated a fractional regression model for recovery rates using the explanatory variables exposed in Section 2. To obtain the maximum likelihood solution, we defined reference groups for the categorical variables: “junior unsecured bonds” for debt instrument, “unsecured debt” for collateral, and the “Telecommunications sector” for industry type. Furthermore, to compare the importance of covariates they must have the same scale. Because all continuous variables are measured as a proportion of total debt, they are bounded to $[0,1]$. Therefore, they have the same scale of the dummy variables that codify the levels of the categorical variables. Debt rank is the only variable with a different scale, since it is measured in an integer scale ranging from 1 to 7. Therefore, we divided all rankings by 7.

Figure 7 shows the average partial effects given by the fractional regression model. With the exception of the debt cushion and the “inventory, accounts receivable and cash” collateral, the most important variables are industry sectors. In particular, the natural products sector, which has the highest recovery rate among all industry sectors, is the most important variable.

3.2 Decision trees

Decision trees are non-parametric models where the data are recursively partitioned into smaller subsets through a sequence of if-then-else conditions on the covariates. The algorithm begins with a “root node” containing all observations. We want to find a regressor, X_j , and a cut-off value on that regressor, K , such that all observations satisfying $X_j \geq K$ go to one “child node”, while all observations satisfying $X_j < K$ go to another child node.

How are the regressor and cut-off-value chosen? Let $S \equiv \sum_i (Y_i - \bar{Y})^2$ denote the total variation of Y in a node. The optimal regressor and cut-off-value are those that maximize the reduction on the variation of Y with respect the parent node,

$$\text{maximize } S_P - S_{C1} - S_{C2}, \quad (5)$$

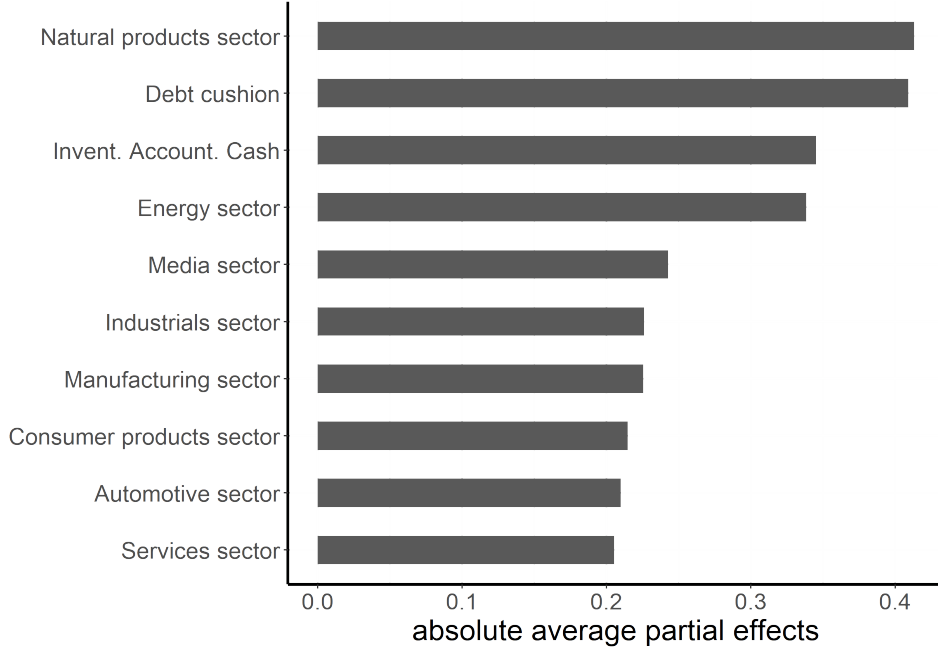


Figure 7: Ten most important determinants of recovery rates given by a fractional regression model. Variable importance is measured by the average partial effects.

where the subscripts P , $C1$ and $C2$ refer to the parent node and its child nodes, respectively. This procedure is then repeated recursively for new nodes, until a minimum allowed number of observations in a node is achieved or the tree reaches a certain size. Important variables will eventually be used in several splits. The unsplit terminal nodes are denoted by “leaves”. A new observation will follow a path through the tree according to its regressor values, and end its trajectory in a leaf. The prediction \hat{Y} for that observation is the sample mean of the training observations in that leaf. This recursive algorithm can easily generate very large trees that will overfit the training data, and have poor accuracy on new data. To counterbalance this, there are several tree “pruning” algorithms that penalize large number of leaves in a tree (see, e.g. Hastie et al., 2009).

Like most machine learning algorithms, decision trees have a set of “hyper-parameters” that need to be tuned. These are parameters that are not learned when training the model. The three hyper-parameters to be tuned are the minimum number of observations in a node in order for a split to be attempted, the maximum depth of the tree, and a complexity parameter that determines how aggressively we prune the tree. To obtain the optimal hyper-parameters we estimated trees using all possible combinations of common hyper-parameter values. The best set of hyper-parameters is the one that generates the tree with lowest out-of-sample mean squared error given by a 10-fold cross-validation.

Figure 8 illustrates a decision tree for predicting Moody’s URD recovery rates. For clarity of illustration we have deliberately fit a small and shallow tree. Nevertheless, regardless of the chosen tree depth, the nodes that are split at a given level always use the same variables and

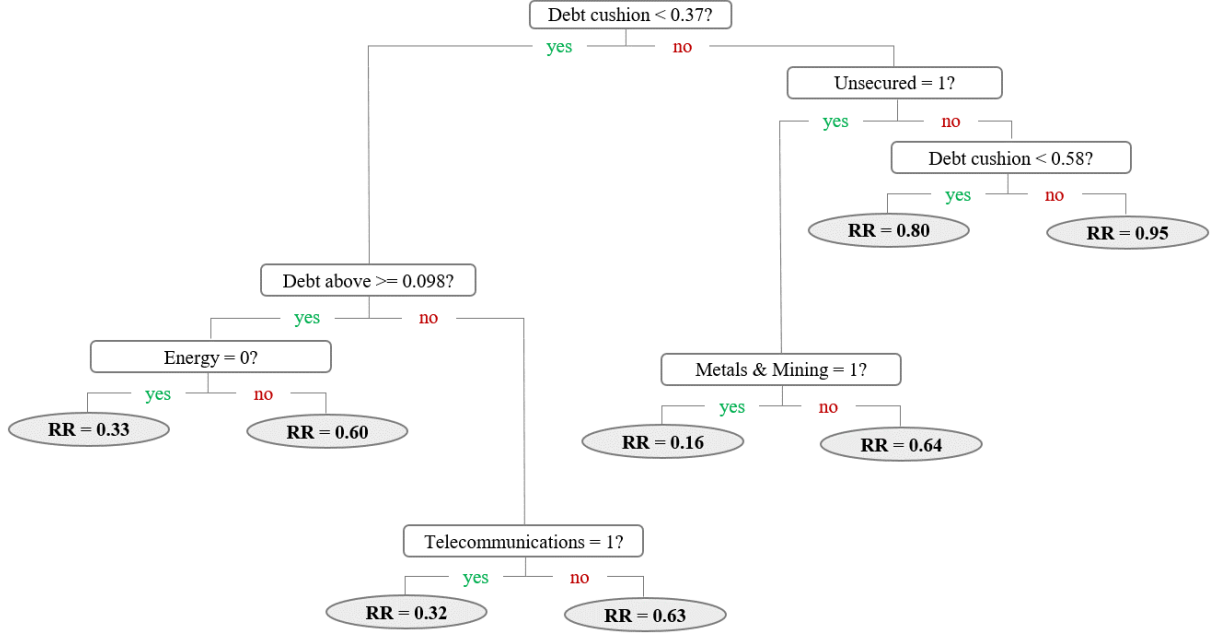


Figure 8: Decision tree to predict recovery rates from Moody’s URD recoveries. For clarity of illustration, a small tree was deliberately fit to the data.

cut-off values. For a given observation, first it is asked if the debt cushion is less than 0.37. If the answer is “no”, it is asked whether the debt is unsecured. If this is not the case, it is asked if the debt cushion is less than 0.58. If the answer is again “no”, then the predicted recovery rate for this observation is 0.95. Naturally, the same reasoning applies to any path followed by an observation along the other branches. This model is fairly interpretable – to some extent we can understand which variables play a role in predicting recoveries, and the direction of their effects on recoveries. However, very large trees are often generated. For example, one branch in the optimal tree for predicting recoveries in our dataset asks 16 questions about the regressors before reaching the leaf. In this case, understanding which covariates are important is not straightforward.

Decision trees are good at ignoring redundant variables. In fact, we could measure variable importance by counting the number of times each regressor was involved in the binary splits. For instance, the debt cushion is used twice in the tree shown in Figure 8, and therefore it must be important. But this would ignore the fact that splits near the root node achieve a greater reduction in the variation of recovery rates than splits near the leaves. A more meaningful measure of importance is the sum of reductions in variation achieved in all splits where a variable participated. Figure 9 shows the cumulative reduction in variation for the 10 most important variables according to this measure. The values were normalized such that the most important variable gets a score of 100. The contrast between the most important predictors given by the decision tree and the fractional regression model (Figure 7) is evident. The decision tree highlights the importance of the debt position in terms of claim priority in the firm’s liabilities, which is measured by the debt cushion, debt above, and ranking. The type

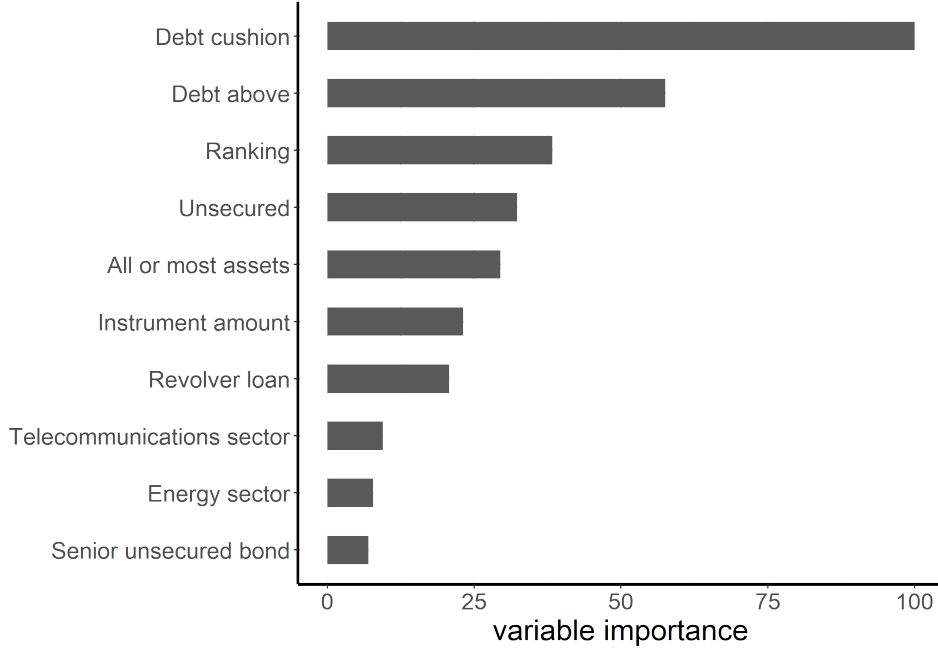


Figure 9: The ten most important determinants of recovery rates according to a decision tree model. Variable importance is measured by the cumulative reduction in the variation of recovery rates achieved by a given variable. These measures were normalized such that the most important variable gets a score of 100.

of collateral is also important (“Unsecured” and “All or most assets”). The type of exposure is less important. Furthermore, the firm’s industry sector is not as important as the fractional regression model suggests.

3.3 Gradient boosting machine

The black-box model is a “gradient boosting machine” (Friedman, 2001). This model can be found on the lower-right corner of Figure 1. Boosting machines combine several base models to produce a powerful “committee”. Typically, the base models are decision trees. Therefore, the prediction \hat{Y}_i for a given observation will be the sum of the predictions given by a set of K decision trees $\{f_k\}_{k=1}^K$,

$$\hat{Y}_i = \sum_{k=1}^K f_k(\mathbf{X}_i; \mathbf{w}_k), \quad (6)$$

where \mathbf{w}_k is a set of leaf “weights” for the k th tree.

For the optimization problem to be computationally tractable, trees are added to the committee sequentially. The first decision tree, $f_1(\mathbf{X})$, is a regular decision tree, as described in subsection 3.2. The next decision trees, $\{f_k(\mathbf{X})\}_{k=2}^K$, are added sequentially, and do not change the structure of those that have already been added. They are incremental trees (“boosts”)

that minimize the following loss function over the estimation sample:

$$\sum_{i=1}^n L\left(Y_i, \hat{Y}_i^{(k-1)} + f_k(\mathbf{X}_i)\right) + \gamma T + \frac{1}{2}\lambda \|\mathbf{w}_k\|^2. \quad (7)$$

The first term is a squared-error loss,

$$L\left(Y_i, \hat{Y}_i^{(k-1)} + f_k(\mathbf{X}_i)\right) = \left(Y_i - \hat{Y}_i^{(k-1)} - f_k(\mathbf{X}_i)\right)^2 = (\hat{\varepsilon}_i - f_k(\mathbf{X}_i))^2, \quad (8)$$

where $\hat{\varepsilon}_i$ is the residual of the previous tree for the i th observation. Therefore, this term selects the tree that best fits the current residuals as the one to be added to the committee. The last two terms are regularization terms that penalize complex trees in order to prevent the committee from overfitting the data. The parameter γ is a penalization term on the number of terminal nodes, T , and λ is a penalization term on the magnitude of the weights. A gradient descent algorithm is used to minimize the loss function when adding new trees.

We use an efficient implementation of this optimization problem known as eXtreme Gradient Boosting, or XGBoost (Chen and Guestrin, 2016). This is probably the best “off-the-shelf” algorithm for a wide range of predictive tasks: about 60% of the winning solutions posted on Kaggle during 2015, and all the top 10 solutions in the KDD Cup 2015 used XGBoost. We optimized three hyper-parameters: the maximum depth of the trees in the committee, the number of trees in the committee, and the “learning rate” of the gradient descent algorithm used in the minimization of the loss function in Equation 7. Again, we look at all possible combinations of common hyper-parameter values, and the best set of hyper-parameters is the one that generates the committee with lowest out-of-sample mean squared error given by a 10-fold cross-validation. We have found that the best committee contains 700 trees, each with a maximum depth from the root to the leaves of 8 splits.

3.4 Predictive accuracy

Table 2 compares the out-of-sample accuracy for predicting recoveries given by the three models. The first column shows the out-of-sample mean squared error obtained using a 10-fold cross validation. The parametric fractional regression model gives the worst predictions on average. The non-parametric decision tree has slightly lower error than the fractional regression (12% lower). As expected, the most accurate model is the gradient boosting machine, with an out-of-sample mean squared error 43% lower than the fractional regression, and 35% lower than the decision tree.

Kalotay and Altman (2017) note that k -fold cross-validation may result in different results from the same borrower ending up in both the estimation and evaluation data sets. This may give an over-optimistic predictive accuracy because the estimation and validation sets are not truly independent. While this may not be an issue in large data sets such as Moody’s URD, we can easily estimate “out-of-time” accuracy errors by dividing the data into two sets according to the date of default. This guarantees that exposures from the same borrower are in the same

Model	MSE (10-fold cross validation)	MSE (out-of-time)
Fractional regression model	0.084	0.202
Decision tree	0.074	0.112
Gradient boosting machine	0.048	0.105

Table 2: Out-of-sample mean squared errors given by 10-fold cross-validation and out-of-time validation for Moody’s URD recovery rates.

sample. Furthermore, we can ensure that observations used for estimation occurred before those used for validation, mimicking the actual experience of a bank. Table 2 also shows the out-of-sample accuracy for the three models using out-of-time validation. The estimation sample contains defaults that occurred between 1987 and 2005, corresponding to about 80% of the data, while the evaluation sample contains defaults that occurred between 2006 and 2010. The ordering of the models in terms of accuracy is the same as that obtained using the 10-fold cross-validation.

4 Interpreting the black-box

The black-box model is clearly better than the other two in terms of predictive accuracy. Of course, by itself the black-box model does not provide any knowledge of its inner workings, since it consists of a large ensemble of 700 complex decision trees. In this section, we show how to look inside it.

4.1 Regressor permutation

The most simple way to measure the importance of a regressor for a black-box model is to randomly permute its values, and evaluate the change in the model prediction error. The random permutation breaks the relationship between the covariates and the output. A regressor is “important” if, after shuffling its values, the model error increases considerably, and “unimportant” if the model error does not change significantly. If the error increases, then the model relied on the regressor in question to generate its predictions. On the other hand, if the error does not change, the model ignored that regressor. Breiman (2001) introduced this approach for measuring variable importance in random forests. However, this technique is model-agnostic and can be applied to any black-box model.

Figure 10 shows the ten most important determinants of recovery rates according to the gradient boosting machine. Increments in model error due to variable permutation are measured by mean squared errors. For the gradient boosting machine, the two most important variables are related to the relative seniority of the debt in the firm’s liabilities. However, in contrast with the decision tree, the debt ranking is missing from the list of the 10 most important variables. Instead, the relative debt amount with respect to the firm’s total debt occupies the

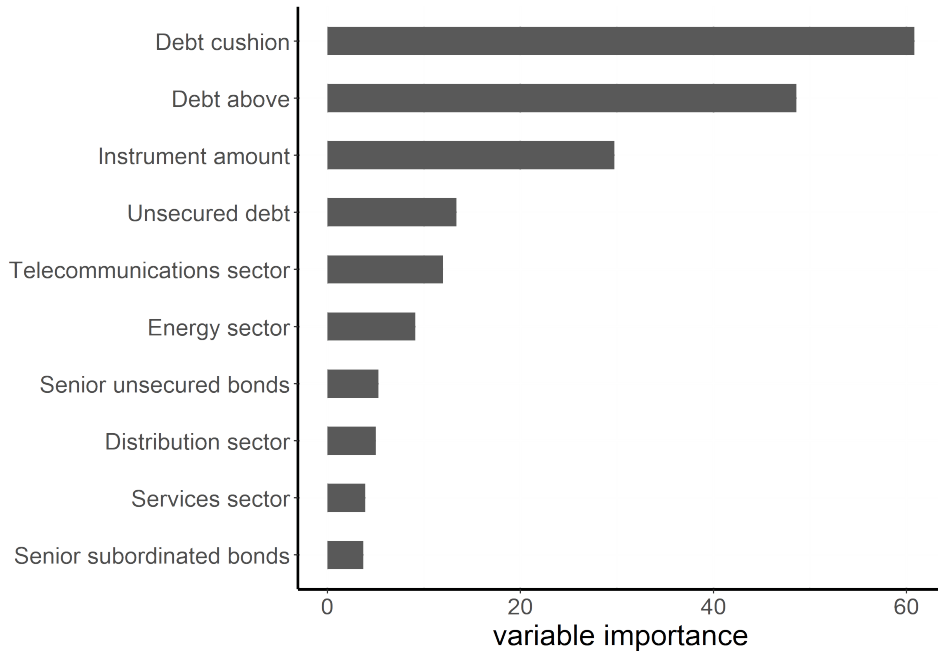


Figure 10: The ten most important determinants of recovery rates according to a gradient boosting machine. Variable importance is measured by the change in the model accuracy when regressor values are randomly permuted.

third position. Unsecured debt is the fourth most important determinant of recovery rates. It is followed by industry sectors and types of debt instrument (senior unsecured bonds and senior subordinated bonds). The contrast with the most important determinants of recoveries given by the fractional regression (Figure 7) is even greater.

Permuting the values of a regressor also destroys the relationship with other covariates. Therefore, importance measures based on permutation also take into account the effect of variable interactions to the model predictions. This may be a disadvantage of this method, since the strength of the interaction between two variables contributes to the importance measures of both variables. The following technique is robust to this issue.

4.2 Shapley values

Shapley values (Lundberg and Lee, 2017) are based on cooperative game theory, and provide one possible answer to the following problem: a coalition of players cooperates and obtains a certain payout from the cooperation; however, some players may contribute more to the total payout than others; how to fairly distribute the payout among the players in any particular game?

When this problem is applied to regression analysis: the game is predicting Y for an observation, the players are the regressors that collaborate to receive the final payout, the importance of a regressor is measured by how much it contributes to the prediction, and the final payout is the prediction minus the average prediction for all observations.

Let $\mathbf{X} = \{X_1, X_2, \dots, X_p\}$ denote the set of p covariates. Let $\mathbf{X}_{\setminus j}$ denote the subset of \mathbf{X} that excludes regressor X_j , that is $\mathbf{X}_{\setminus j} \equiv \mathbf{X} \setminus X_j$, and let S denote all possible subsets of $\mathbf{X}_{\setminus j}$. For instance, if we have $p = 3$ regressors, $\mathbf{X} = \{X_1, X_2, X_3\}$, and we exclude regressor X_1 from \mathbf{X} , then $S = \{\emptyset, X_2, X_3, \{X_2, X_3\}\}$. The Shapley value for X_j is a weighted sum of its marginal contribution to a prediction over all possible coalitions that exclude it:

$$\phi(X_j) = \sum_{S \subseteq \mathbf{X}_{\setminus j}} \frac{|S|!(p-1-|S|)!}{p!} [f_{S \cup X_j}(S \cup X_j) - f_S(S)]. \quad (9)$$

The global importance of variable X_j is given by the sum of the absolute Shapley values for all observations in the data:

$$I_j = \sum_{i=1}^n |\phi_i(X_j)| \quad (10)$$

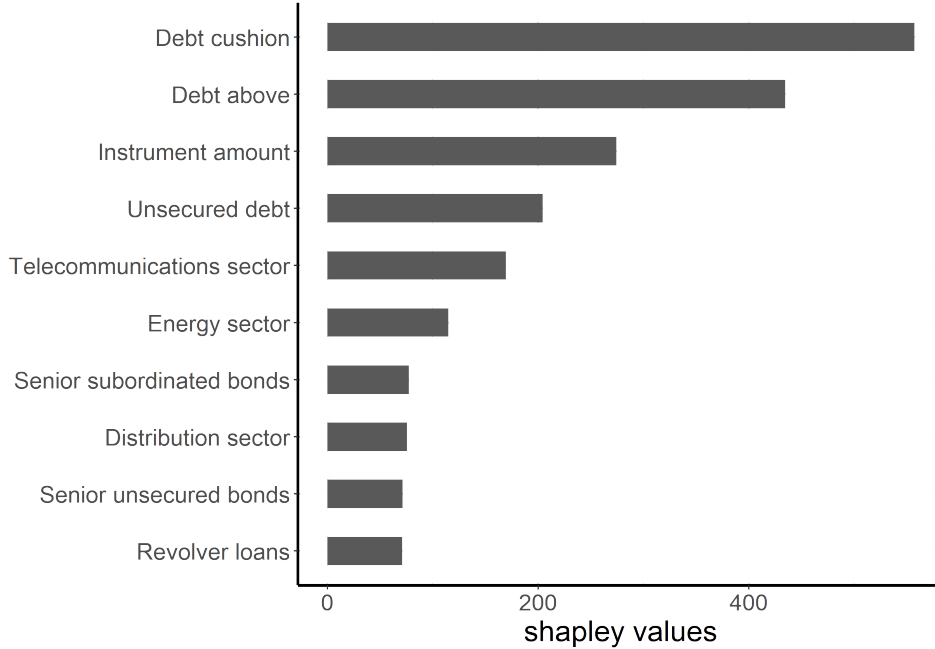


Figure 11: The ten most important determinants of recovery rates according to a gradient boosting machine. Variable importance is measured by Shapley values.

Figure 11 shows the 10 highest Shapley values given by the gradient boosting machine. Comparing figures 10 and 11, we can note that the two methods for explaining the gradient boosting machine show a strong consensus on which determinants are the most important. The only differences are observed in the last positions of the top 10 ranking, affecting industry sectors and instrument types.

4.3 Accumulated local effects

Variable permutations and Shapley values just rank the determinants of recovery rates in terms of their importance. However, we may be interested in assessing whether the relationship between recovery rates and a given regressor is positive or negative, linear or non-linear, convex or

concave. An accumulated local effects (ALE) plot (Apley and Zhu, 2020) is a visual representation of how a regressor influences the model predictions. It indicates the type of relationship between the dependent variable and the regressors. For example, if the true relationship is linear, such plot will actually show a linear dependence. An ALE plot does not give the usual *ceteris paribus* effect prevalent in applied econometrics. It shows how the model output varies as a function of a regressor, averaging the effects of other covariates.

For a given regressor, X_j , we first divide its range using a grid with K bins. The limits of the grid are indexed by $k = 0, 1, \dots, K$. Let $\{Z_k\}_{k=0}^K$ denote the set of values that define the grid. Typically, the Z_k are chosen as the (k/K) -quantiles of the empirical distribution of X_j , with Z_0 chosen just below the smallest observation, and Z_K equal to the largest observation.

Denote the set of observations with $Z_{k-1} < X_j \leq Z_k$ by S_k , and the number of observations in each bin by n_k , with $k = 1, \dots, K$. Finally, let $k(X_j)$ denote the index of the bin where a given value of X_j falls. The accumulated local effect is

$$\text{ALE}(X_j) = \sum_{k=1}^{k(X_j)} \frac{1}{n_k} \sum_{i \in S_k} [f(Z_k, \mathbf{X}_{\setminus j, i}) - f(Z_{k-1}, \mathbf{X}_{\setminus j, i})]. \quad (11)$$

The innermost sum loops over all observations in a given bin. For each of these observations, we obtain the difference between the model predictions with X_j equal to the upper limit of the bin, Z_k , and X_j equal to the lower limit of the bin, Z_{k-1} . We divide this sum by the number of observations in that bin, n_k , to obtain the average local effect of X_j on the model output. The outermost sum accumulates the local average effects up to a given value of X_j . The plot of $\text{ALE}(X_j)$ as a function of X_j provides a visualization of the dependence of recoveries on X_j across its range. The accumulated local effects are typically centered so that the mean effect is zero,

$$\text{ALE}(X_j) \leftarrow \text{ALE}(X_j) - \frac{1}{n} \sum_{i=1}^n \text{ALE}(X_{ji}). \quad (12)$$

Figure 12 shows the accumulated local effects plots for the four main determinants of recovery rates according to the gradient boosting machine. Three of those are numeric variables, and one is a dummy variable. To eliminate statistical artifacts we applied a loess smoother to the effects of numeric variables. The gray bands represent the 95% confidence interval of the loess smoother. As anticipated, the recovery rate is lowest when the debt junior to a given exposure is small, and monotonically increases as the debt cushion grows. The opposite effect is observed for the proportion of senior debt in terms of claim priority – lower values of debt above result in higher recoveries. However, we can see a plateau for mid-range values of debt above. The effect of debt being unsecured by collateral is the expected – secured debts have higher recoveries. The most interesting effect is that of the outstanding amount at default relative to the obligor’s total debt. When the debt amount has a low weight in the obligor’s liabilities the recovery is higher. Then it declines as this proportion increases. When the instrument amount reaches about 30% recoveries achieve a minimum. Then for instrument amounts greater than 30%,

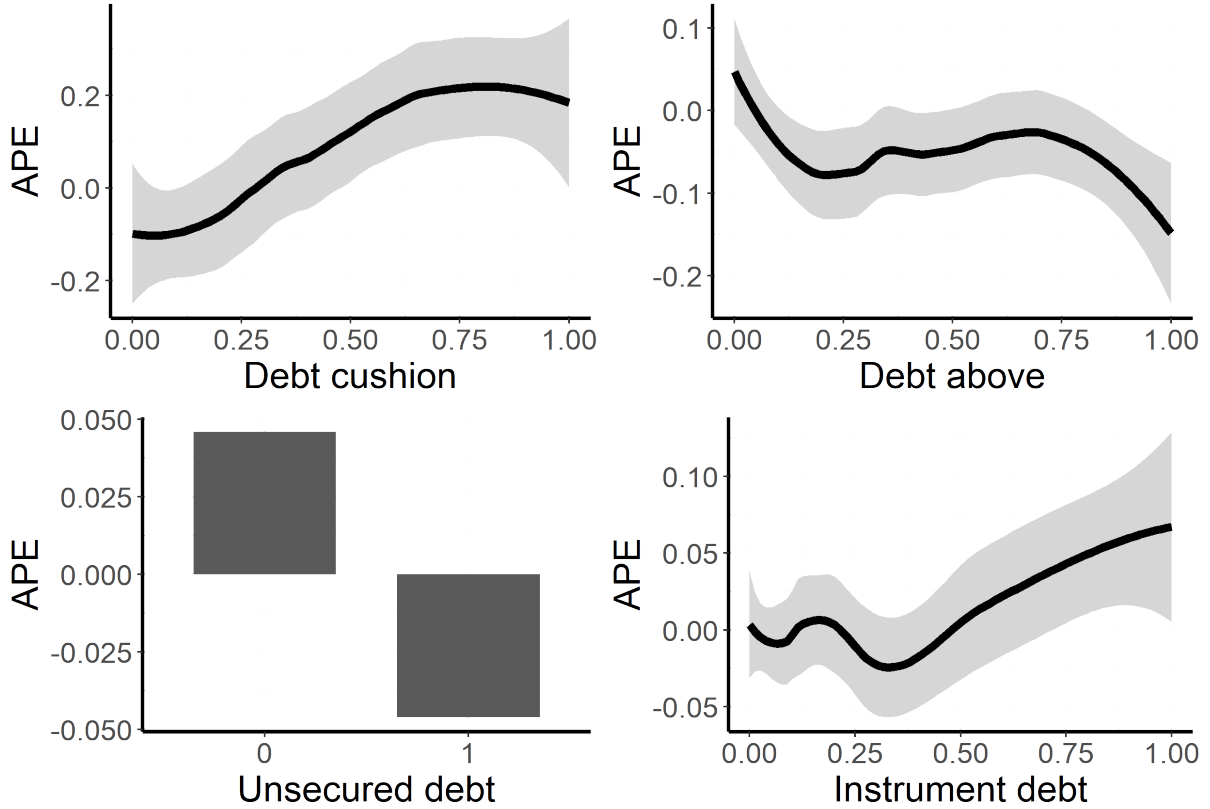


Figure 12: Accumulated local effects plots for the four main determinants of recovery rates according to a gradient boosting machine. A loess smoother was applied to the effects of numeric variables (dark line). The gray band shows the 95% confidence interval of the loess smoother.

recoveries increase monotonically as this variable increases. Neither the univariate analysis nor the fractional regression capture this effect.

5 Conclusions

In this paper, we show that banks do not have to sacrifice predictive accuracy at the cost of model transparency to be compliant with the regulatory requirements on the Basel agreements. We demonstrate this by showing that the predictions of recovery rates given by a black-box model – the gradient boosting machine – can be easily explained in terms of their inputs. We do so by using two techniques to rank the determinants of recovery rates in terms of their importance – variable permutations and Shapley values –, and a technique for assessing the nature of the relationship between recovery rates and the covariates – accumulated local effects plots.

We show that the most important determinants of recovery rates according to the black-box are quite different from those given by the parametric glass-box model. Because black-box models fit better the data, banks should consider the determinants of recoveries suggested

by these models in lending decisions and pricing of credit exposures. Of course, glass-box models allow an immediate understanding of the impact of each covariate. Explaining the model outcomes does not involve any additional effort. On the other hand, black-box models require the additional effort of “X-raying” the box. For instance, the calculation of Shapley values can be computationally intensive when the number of covariates is large. Nevertheless, the considerable differences observed in predictive accuracy certainly justify this endeavor.

References

- Acharya, V.V., Bharath, S.T., Srinivasan, A., 2007. Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics* 85, 787–821.
- Altman, E.I. and Kishore, V.M., 1996. Almost everything you wanted to know about recoveries on defaulted bonds, *Financial Analysts Journal* (November–December).
- Aple, D.W., Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B* 82, 1059–1086
- Bastos, J.A., 2010. Forecasting bank loans loss-given-default. *Journal of Banking & Finance* 34, 2510–2517.
- Bastos, J.A., 2014. Ensemble predictions of recovery rates. *Journal of Financial Services Research* 46, 177–193.
- Basel Committee on Banking Supervision, 2006. International convergence of capital measurement and capital standards.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1983. *Classification and Regression Trees*. Wadsworth, Belmont CA.
- Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- Chen, T., Guestrin, E., 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Davidenko, S.A., Franks, J., 2008. Do bankruptcy codes matter? A study of defaults in France, Germany, and the U.K. *Journal of Finance* 63, 565–608.
- Dermine, J., and Neto de Carvalho, C., 2006. Bank loan losses-given-default: A case study. *Journal of Banking & Finance* 30, 1219–1243.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29, 1189–1232.

- Freund, Y. and Schapire, R., 1996. Experiments with a new boosting algorithm. In *Machine Learning: Proceedings of the Thirteenth International Conference* 148–156. Morgan Kaufman, San Francisco.
- Hastie, T., Tibshirani, R., 1990. *Generalized Additive Models*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability.
- Hastie, T., Tibshirani, R., Friedman, J.H., 2009. *The Elements of Statistical Learning*, 2nd ed. Springer-Verlag.
- Gupton, G. Stein, R., 2005. LossCalc V2: Dynamic prediction of LGD – Modeling methodology. Moody’s KMV.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2014. *An Introduction to Statistical Learning*. Springer, New York.
- Kalotay, E.A., Altman, E.I., 2017. Intertemporal forecasts of defaulted bond recoveries and portfolio losses. *Review of Finance* 21, 433–463.
- Loterman, G., Brown, I., Martens, D., Mues, C., Baesens, B., 2012. Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting* 28, 161–170.
- Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. *NIPS’17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.
- Papke, L. E. and Wooldridge, J. M., 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 11, 619–632.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, 58, 267–288.
- Vapnik, V.N., 1995. *The Nature of Statistical Learning*. Springer, New York.
- Varma, P., Cantor, R., 2005. Determinants of recovery rates on defaulted bonds and loans for North American corporate issuers: 1983–2003. *Journal of Fixed Income* 14, 29–44.